METHOD AND DEVICE FOR VOICE SEPARATION OF COMPOUND VOICE DATA, METHOD AND DEVICE FOR SPECIFYING SPEAKER, COMPUTER PROGRAM, AND RECORDING MEDIUM

Publication number: JP2003005790 (A)

Also published as:

Publication date: 2003-01-08 Inventor(s): YAMAMOT(

YAMAMOTO TAKAYOSHI +

DJP3364487 (B2)

Applicant(s):

YAMAMOTO TAKAYOSHI; URATA TAKAYUKI +

Classification:

- International: G10I 11/00: G10I

G10L11/00; G10L15/00; G10L15/02; G10L15/20; G10L17/00; G10L21/02; G10L11/00; G10L15/00; G10L17/00; G10L21/00; (IPC1-7): G10L11/00; G10L15/02; G10L15/20; G10L17/00;

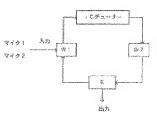
G10L21/02

- European:

Application number: JP20010191289 20010625 Priority number(s): JP20010191289 20010625

Abstract of JP 2003005790 (A)

PROBLEM TO BE SOLVED: To provide a method and a device for separating compound voice data where voice data of several speakers mixedly exist into the voice of every speaker and to provide a method and a device for accurately and quickly specifying the speaker of each separated voice data, SOLUTION: The method for separating compound voice data where voice data of several speakers mixedly exist into the voice data of every speaker has a step (1) where correlation elimination processing is performed to eliminate correlation between the compound voice data and a step (2) where independent component separation processing is performed to separate data subjected to the correlation elimination processing into independent components.



Data supplied from the espacenet database - Worldwide

(S1) Int.CL1

G10L 17/00

11/00

(18) H木岡特許 (JP) (12) 公開特許公報 (A)

線別記号

FΙ

G10L 3/00

9/(0)

(11)特許出繼公開番号 特開2003-5790 (P2003-5790A)

(43)公開日 平成15年1月8日(2003, 1, 5)

テーマコート*(参考)

545A 5D015

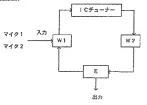
***	•			4, 20		, ,	
15/02 15/20			9/02 9/08		A A		
21/02		9/02			301A		
	答查的 2	段 有	輸収	頁の数21	OL	(全 32 頁)	最終質に続く
(21) 出願番号	特額2001-191289(P2001-191289)	(71	出職人	500262	511		
				山本	隆義		
(22)出版日	平成13年6月25日(2001.6.25)			広島市	ME	要上2丁目54-	-20
		(7)	出職人	501253	822		
				淮田	鄉力		
		ĺ				中年前100年	※1丁目3-5
				- 8	A-4 -44 - 1 - 1 - 1	9C 02277 BACH 4 FOX	,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
		699	発明者	ilisk	165.201		
	44	02.	209330			actorization of	mers on
		広島県広島市西区ご斐上2丁目54-20 (74)代理人 100071283					
		(14)	4.Citi V				
						鍵維 (外:	3 45)
		F5	F ターム(参考) 5D015 AAD2 CC04 CCD5				

(54) 【発明の名称】 複合音声データの音声分離方法、発音者特定方法、複合音声データの音声分離装置、発音者特定 装置、コンピュータプログラム、及び、記録媒体

(57) 【要約】

【課題】 複数の発言者の音声データが混在する混在音 声データを、発言者毎の音声に分離する方法及び装置。 さらに分離された各斉声データの発音者を特定すること を、正確にかつ高速に行うことができる方法及び装置を 提供する。

【解決手段】 複数発音者の音声データが混在している 混在音声データを、発言者毎の音声データに分離する音 声データ分離方法において、(1) 前記混在音声データ を互いに無相關化するための無相關化処理を行うステッ プと、(2) 輪記無相関化処理の行われたデータを独立 成分に分離するための独立成分分離処理を行うステップ とを有する。



【特許請求の範囲】

【請求項1】 複数発言者の音声データが混在している 混在音声データを、発言者毎の音声データに分離する音 声データ分離方法において、(1) 前記混在音声データ を互いに無相関化するための無相関化処理を行うステッ プと、(2) 前記無相関化処理の行われたデータを独立 成分に分離するための独立成分分離処理を行うステップ と、を有することを特徴とする音声分離方法。

【請求項2】 請求項1に記載の音声分離方法におい て、

前記独立成分分離の行われたデータの分離性が不十分な 場合には、分離性が十分になるまで、前記独立成分分離 処理の行われたデータについて、前記無相関化処理及び 前記独立成分分離処理を繰り返し行うことを特徴とする 音雷分離方法。

【請求項3】 請求項1又は請求項2に記載の音声分離 方法において、

前記独立成分分離処理として、非ガウス性のデータを独 立成分に分離するための非ガウス性独立成分分離処理 常性独立成分分離処理と、有色性のデータを独立成分に 分離するための有色性独立成分分離処理とを準備し、デ 一夕の性質により、前記非ガウス性独立成分分離処理、 前記非定常性独立成分分離処理、及び、前記有色性独立 成分分離処理のうちのいずれかの処理を行うことを特徴 とする音声分離方法。

【請求項4】 請求項3に記載の音声分離方法におい

最初に行われる独立成分分離処理は、非ガウス性のデー 処理であることを特徴とする音声分離方法。

【請求項5】 請求項1乃至請求項4に記載の音声分離 方法において

前記無相關化処理は、少なくとも主成分分析及78因子分 析を行うことを特徴とする音声分離方法。

【請求項6】 複数発言者の音声データが混在している 混在音声データを、発言者毎の音声データに分離し、該 発言者毎の音声データにつき発言者を特定する発言者特 定方法において、(1)請求項1乃至請求項5のいずれ タが混在している混在音声データを、発言者毎の音声デ ータに分離するステップと、(2)発言者毎に該発言者 を特定するための特定パラメータを準備するステップ と、(3)分離された前記発言者毎の音声データにつ き、前型特定パラメータを参照して、発言者を特定する ステップと、を有することを特徴とする発言者特定方 法。

【蕭求項7】 蕭求項6に記載の発言者特定方法におい 前記特定パラメータは、発言者が母音を発音した際のホ 50 発言者特定方法において、

ルマント周波数であり、 分離された前記発言者毎の音声データにつき、ホルマン ト層波数を求め、求められたホルマント層波数に関し

て、前記特定パラメータとしてのホルマント周波数を参 照して、発言者を特定することを特徴とする発言者特定 方法。

【請求項8】 請求項7に記載の発言者特定方法におい て、

前記特定パラメータは、発言者が母音を発音した際の第 10 1ホルマント周波数及び第2ホルマント周波数であり、 分離された前記発言者毎の音声データにつき、第1ホル マント周波数及75年2ホルマント周波数を求め、求めら れた第1ホルマント周波数及び第2ホルマント周波数に 関して、前記特定パラメータとしての第1ホルマント周 波数及び第2ホルマント周波数を参照して、発言者を特 定することを特徴とする発言者特定方法。

【請求項9】 請求項6乃至請求項8のいずれかに記載 の発言者特定方法において、

分離された前記発言者毎の音声データにつき、前記特定 と、非定常性のデータを独立成分に分離するための非定 20 パラメータを参照して発言者を特定するステップにて発 言者を特定できなかった場合には、

> 該音声データから複数の時点のホルマント周波数を求 め、求められた複数時占のホルマント周波数に関して、 前記特定パラメータとしての複数時点のホルマント圏波 数を参照して、発言者を特定することを特徴とする発言 老特定方法。

> 【請求項10】 請求項6乃至請求項9のいずれかに記 載の発言者特定方法において、

分離された前記発言者毎の音声データにつき、前記特定 タを独立成分に分離するための非ガウス件独立成分分離 30 パラメータを参照して発言者を特定するステップにて発 言者を特定できなかった場合には、

> 該音声データから有声音データを分離し、 該有声音デー タにつき、ホルマント周波数を求め、求められたホルマ ント最波数に関して、前記特定パラメータとしてのホル マント周波数を参照して、発言者を特定することを特徴 とする発言者特定方法。

【請求項11】 請求項10に記載の発言者特定方法に おいて、

分離された前記発言者毎の音声データにつき、前記特定 かに記載の音声分離方法により、複数発言者の音声デー 40 パラメータを参照して発音者を特定するステップにて発 言者を特定できなかった場合には、

> 該音声データから有声音データを分離し、該有声音デー タにつき。第1ホルマント間波数及び第2ホルマント周 波数を求め、求められた第1ホルマント周波数及び第2 ホルマント層波数に関して、前記特定パラメータとして の第1ホルマント周波数及び第2ホルマント周波数を参 照して、発言者を特定することを特徴とする発言者特定 方法。

【請求項12】 請求項10または請求項11に影散の

分離された訴訟発言者毎の音声テータにつき、前記特定 パラメータを参照して発言者を特定するステップにて発 言者を特定できなかった場合には、

総有声音データにつき、複数の時点のホルマント周波数 を求め、求められた複数時点のホルマント周波数に関し て、前記特定パラメータとしての複数時点のホルマント 周波数を参照して、発音者を特定することを特徴とする 発音者特定方法。

【請求項13】 請求項10又は請求項11に記載の発 営者特定方法において。

前記音声データから前記有声音データを分離する際に、 該音声データに対して独立成分に分離するための独立成 分分離処理が行われることを特徴とする発言者特定方 法。

【請求項14】 複数発言者の音声データが混在している混在音声データから、議事録を作成する議事録作成方法において、

請求項6万至請求項13のいずれかに記載の発言者特定 万法により、分離された前記発言者毎の音声データにつ き、発言者を特定するステップと、

特定された発営者と、該発営者の発言とを対応付けて記 録媒体に出わすることにより、線事線を作成するステッ プと、を有することを特徴とする議事録作成方法。

【輸求項15】 複数発言者の畜声データが泥在してい る混在音声データを、発営者毎の音声データに分離する 音声データ分離装置において.

前記混在音声データを互いに無相関化するために無相関 化処理を行い、

輸記無規則化処理の行われたデータを独立成分に分離するために独立成分分離処理を行うことを特徴とする音声 3 分類性の計能記管温等毎の音声データにつき、第1 ホルケン 分離装施。 マント湯参数及び第2 ホルマンと 層が数を求め、求めら

【請求項16】 請求項15に記載の音声分離装置において。

輸配独立成分分離の行われたデータの分離性が不十分な 場合には、分離性が十分になるまで、輸配独立成分分離 処理の行われたデータについて、前記無相限化処理及び 新配独立成分分離処理を繰り返し行うことを特徴とする 音声分種談頭。

【請求項17】 請求項15又は請求項16に配載の音 声分離装置において、

データの性質により、箱記徳立成分分離処理として、非 ガウス性のデータを独立成分に分離するための非ガウス 整独立成分が概憶理。非定等後のデータを独立成分に分 離するための非定常性独立成分分離処理。有色性のデー タを独立成分に分離するための有色性独立成分分離処理。 のうちのいずれかの処理を行うことを特徴とする音 声分離装置。

【請求項18】 請求項17に記載の音声分離装置において

験初に行われる独立成分分離処理は、非ガウス性のデー 50 は、

タを独立成分に分離するための非ガウス性独立成分分額 処理であることを特額とする音声分離装置。

【錦求項19】 請求項15万至請求項18に記載の音 市分輔装置において、

前記無相關化処理は、少なくとも主成分分析及び因子分析を行うことを特徴とする帝部分難時間。

【輸求項20】 複数発酵者の音声データが混在している混在音声データを、発営者毎の音声データに分離し、

該発言者毎の音声データにつき発言者を特定する発言者 10 特定装置において、 請求項15乃至請求項19のいずれかに記載の音声分離 装置により、複数発言者の書面データが固在している泥

在彦声データを、発宮省毎の落声データに分離し、 分離された前記発言者毎の落声データにつき、発言者毎 に該発言者を特定するための特定パラメータを参照して 発言者を特定することを特徴とするを宣者特定装置。

【請求項21】 請求項20に記載の発言者特定装置に おいて、

前記特定パラメータは、発営者が母音を発音した際のホ 20 ルマント局波数であり、

M ルマント海波数であり、 分離された前記発音者毎の音声データにつき、ホルマン ト周波数を求め、求められたホルマント周波数に関して、前配特定パラメータとしてのホルマント異波数を参

て、前記特定パラメータとしてのホルマント周波数を参 照して、発言者を特定することを特徴とする発音者特定 装騰。

【請求項22】 請求項21に記載の発言者特定装置に おいて、

前記特定パラメータは、発言者が母音を発音した際の第 1 ホルマント周波数及び第2ホルマント周波数であり、

分離された前記院沿着毎の溶南データにつき、第1 ホル マント周波数及び第2 ホルマント周波数を求め、求めら れた第1 ホルマント周波数及び第2 ホルマント周波数に 関して、前記特定パラメータとしての第1 ホルマント周 波数反び第2 ホルマント周波数を参照して、発音者を持 定することを特徴とする発音者を近装選

【請求項23】 請求項20万至衡求項22のいずれか に記載の発言者特定装置において、

分離された前紀発言者毎の音声データにつき、前記特定 パラメータを参照して発客者を特定できなかった場合に 40 は、

裁音声データから複数の時点のホルマント開設数を求 め、求められた複数時点のホルマント開設数に関して、 前記時定パラメータとしての複数時点のホルマント開設 数を参照して、発言者を特定することを特徴とする発音 送籍字葉電

【請求項24】 請求項20乃至請求項23のいずれか に記載の発言者特定装置において、

分離された前記発言者毎の音声データにつき、前記特定 パラメータを参照して発言者を特定できなかった場合に 該音声データから有声音データを分離し、該有声音デー タにつき、ホルマント側波数を求め、求められたホルマ ント周接数に関して、前記特定パラメータとしてのホル マント周波数を参照して、発言者を特定することを特徴 とする発言者特定設置。

【結束項25】 結束項24に記載の発言者特定装置に おいて、

分離された前紀発音者係の音遊データにつき、前紀縁定 パラメータを参照して発音者を特定できなかった場合に

該音声データから有声音データを分離し、該有声音デー タにつき、第1ホルマント周波数及び第2ホルマント眉 波数を求め、求められた第1ホルマント周波数及び第2 ホルマント周波数に関して、前記特定パラメータとして の第1ホルマント周波数及び第2ホルマント局波数を参 照して、発言者を特定することを特徴とする発言者特定 物體。

【請求項26】 請求項24または請求項25に記載の 発言者特定装置において.

パラメータを参照して発言者を特定できなかった場合に ほ.

該右密音データにつき、複数の結古のホルマント開放数 を求め、求められた複数時点のホルマント間接数に関し て、前記特定パラメータとしての複数時点のホルマント 周波数を参照して、発言者を特定することを特徴とする 発育者特定装置。

【請求項27】 請求項24又は請求項25に記載の発 言者特定装置において、

前記音声データから前記有声音データを分離する際に、 該音声データに対して独立成分に分離するための独立成 分分離処理が行われることを特徴とする発言者特定装 激。

【請求項28】 複数発常者の音声データが瀑布してい る潔在音声データから、緻密緑を作成する議事録作成装 徽において、

職求項20乃至継求項27のいずれかに記載の咎言者特 定装置により、分離された前紀発言者毎の音声データに つき、発音者を特定し、

特定された発言者と、影発等者の発言とを対応付けて記 40 緩緩体に出力することにより 鎌事器を作成することを 特徴とする議事銀作成装置。

【継承項29】 離求項1乃至結求項5のいずれかに記 載の音声分離方法を音声分離装置に実行させるためのコ ンピュータプログラム。

【糖求項30】 請求項6万至請求項13のいずれかに 記載の発言者特定方法を発言者特定装置に実行させるた めのコンピュータプログラム。

【請求項31】 請求項29又は請求項30に記載のコ

可能な記録媒体。 【発明の詳細な説明】

[0001]

【発明の縁する技術分野】本発明は、複数発言者の複合 容声データの音声を分離する方法、分離したそれぞれの 音声データの発言者を特定する方法、複数発言者の複合 音声データの音声を分離する装置。分離したそれぞれの 音声データの発言者を特定する装置、コンピュータプロ グラム、及び、記録媒体に関する。

10 [0 0 0 2]

【従来の技術】複数の発営者の音声が混合されて記録さ れている、音声記録媒体中の複合音声データを、発言者 毎に正確に分離する技術が切望されている。具体的に は、複合音声データを、音声の入力と同時進行的に発常 者毎に分離し特定することで、会議の議事録作成を自動 的に行うことのできるような技術が明望されている。

【0003】従来、長時間にわたる会議の議事録を作成 するには、各種の音声記録機器に記録した会議の音声デ 一タを、議事録作成担当者か全て聞きなおし、要約する

分離された前記発言者毎の音声データにつき、前記特定 20 などして議事操を作成していた。この作業は、音声記録 機器の再生と一時停止を何度も繰り返しつつ行う必要が あり、手關と時間がかかる。

> 【0004】また、もう1つの問題は、発言者の特定が 個難であることである。本人が会議に出席した担当者な らまだしも、そうでない担当者が議事録を作成するの は、どの音声がどの発言者によるものなのかを判断する のは非常に困難なことであった。

【0005】従来、混合音声データからの音声分離、発 営者特定に関する技術は幾つか存在してはいるが、1本 30 のマイクに複数人の音声やノイズが混合されて入力され る場合でも分離、特定を正確に行い、さらに、複合音声 の入力と同時進行的に高速な分離・特定処理を行うこと は、時間的に連続な音素データのセグメンテーション、 及び顕音結合の点で非常に難しい課題であった。

【0006】特開2001-27895には、複数の億

等額からの音響信号を分離し、所望の信号を含成出力す るための信号分離方法が記載されている。この発明は、 解析対象の混合音声・音響部号に対し時間・局波数解析 を行い、周波数成分の倍音構成を等る。倍音周波数成分 のうち、立上がり時間及び立下り時間の少なくとも一方 が共通であるか否かで、 同一信号達からの周波数成分で あるかどうかを設定する。その周波数成分を抽出・再構 成することにより、単一倍号源からの倍号を分離する。 【0007】この発明は、混合された循号の相関性や独 立性といった事項を考慮していないので、同じ局或数帯 域に属する混合信号、あるいは同時間帯に存在する混合

【0008】また、特徴2000-97758に記載さ れた音源信号推定装置では、複数の音響信号がそれぞれ ンピュータブログラムを記録したコンピュータ読み取り 50 混在して複数のチャンネルを介して入力されたときに、

信号を分離することは困難である。

各音源信号が設合係数ペクトルと内積減算されて他の音 額信号に加算される混合過程モデルに基づき、混合係数 ベクトルに対抗する分離係物ベクトルを逐次修正しなが ら求め、この分類係数ベクトルを削いて音響信号の推 定、分離を行う(ICAの手法)にあたり、分離係数ペ クトルの逐次修正に用いる修正ペクトルを正規化する高 声信号とそれ以外の領号が相互に混在している信号から それぞれの領号を推定し、分離するに際し、それぞれの 借号パワー変動による推定 分離への影響を軽減するこ とができ、さらに、収棄係数を大きくすることができる 10 性が、時間的・空間的に変動する場合でも、発言者毎の ことから安定かつ高速の係号分離が可能となる、とされ ている。

【0009】この発明は、独立成分解析(1CA)をベ 一スとして分類係数ペクトルを逐次修正しながら行うの で、信号パワーの変動影響を軽減でき、高速分離を実現 するものであるが、様々な信号源からの音源信号はお互 いに独立性を保持しているとは限らない。一般に、たと え独立した信号源からの音源信号であっても混合される と相関件を行してしまっていることが多いが、その点が 表謝されていない。

【0010】また、特簡平9-258788には、総本 間波数の近接した混合音声を適切に区別分離し、音順の 動に制御されず、高品質の分離音商を得ることを目的と した音声分離方法および装置が記載されている。この発 明では、入力音響館号中に含まれる音声信号の有声音部 分と無声音部分の内の有声音部分は有声音の音激方面の 情報を加味しながら個別に抽出し、抽出された有声容部 分を複数の有声音に分化して有声音の群として抽出し、 音声信号の無声音部分は入力音響信号から有声音部分を 滅算して抽出した残差から各有声音の群の無声音に相当 30 する音響信号の成分として抽出し、各別に抽出された有 声音の群に無声音を補充して音声信号を抽出することに よって上記目的を実現する。

【0011】この発明は、音源方位の情報を抽出する音 湖定位部を有しているが、同じ方向から異なる音声が発 せられた場合は分離が困難となる。また、複数の発言者 が同じ出音、あるいは有窓道を築したときはそれらの分 額が困難であると思われる。

[0012]

【経明が解決しようとする振順】以上のような従来技術 40 が存する種々の問題点を解決すべく、本発明は、複数の 発出者の音声データが誕在する混在音声データを、発言 者類の音声に分離する方法及び装置、さらに分離された 各音声データの発音者を特定することを、正確にかつ高 速に行うことができる方法及び装置の提供を主たる目的 とする.

[0013]

【課題を解決するための手段】上記の課題を解決するた めに、本出願に係る第1の発明は、複数発言者の音声デ データに分離する音声データ分離方法において、「11 前部源在音声データを互いに無料關化するための無料器 化処理を行うステップと、(2) 前記無相關化処理の行 われたデータを独立成分に分離するための独立成分分類 処理を行うステップとを有することを特徴とする音声分 離方法である。このような第1の発料によれば、入力さ れる混在音声データ(生データ)に含まれる各音声デー タの料理性、及び独立性の新性質をともに考慮し、複動 の音声データや組入する雑音などの有する相関性や独立 音声に正確に分離することができる。

【0014】また、本出願に係る第2の発明は、第1の 発明である音声分離方法において、前紀独立成分分離の 行われたデータの分離性が不十分な場合には、分離性が 十分になるまで、前記独立成分分離処理の行われたデー タについて、前記無相關化処理及び前記独立成分分離処 期を繰り返し行うことを特徴とする音声分離方法であ る。このような第2の発明によれば、混在音声データを 音潔毎の音声データに充分に分離させることができる。

20 【0015】また、本出額に係る第3の発明は、第1又 は第2の発明である音声分離方法において、前記独立成 分分類処理として、非カウス性のデータを独立成分に分 鎌するための非ガウス性純立成分分離処理と、非定常性 のデータを智力成分に分離するための非定常性執力成分 分離処理と、有色性のデータを独立成分に分離するため の有色性独立成分分離処理とを準備し、データの性質に より、前記非ガウス性独立成分分離処理、前記非定常性 独立成分分離処理,及び,給記有色件独立成分分離処理 のうちのいずれかの処理を行うことを特徴とする音声分 離方法である。このような第3の発明によれば、無相関 化処理の行われたデータの件質に応じて影適な独立成分 分離処理を行うことができるから、混在音声データを音 源毎の音声データにより効果的に分離させることができ Z.,

【0016】また、本出願に係る第4の発明は、第3の 発明である音声分離方法において、最初に行われる独立 成分分離処理は、非ガウス件のデータを独立成分に分離 するための非ガウス性独立成分分離処理であることを特 数とする音声分離方法である。非ガウス性独立成分分離 処理は他の仲立成分分処理方法におってその前処理とし ての無相関化処理の影響を受けやすいから このような 第4の発明によれば 最初に非ガウス性独立成分分離処 理を行うことにより、無料関化钒器がらまく生行された かどうかを、該無相關化処理に引き続く非ガウス骨独立 成分分離処理によって効果的に評価することが可能とな

【0017】また、本出願に係る第5の発明は、第1万 至第4の発明である音声分離方法において、前記無相勝 化処理は、少なくとも主成分分析及び因子分析を行うこ ータが罷在している混在音声データを、発言者毎の音声 50 とを特徴とする音声分離方法である。このような第5の

発明によれば、各主成分の寄与率を求めて累積寄与率が 所定のしきい値を越えるところの成分数を次数とするこ となどにより、採用する主成分データの数(次数)を決 定した上で、効果的に無相関化処理を行うことが可能と to 3.

【0018】また、本出願に係る第6の発明は、複数発 言者の音声データが混在している混在音声データを、発 言者毎の音声データに分離し、該発言者毎の音声データ につき発言者を特定する発言者特定方法において、

(1)第1乃第5のいずれかの発明の音声分離方法によ 10 り、複数発言者の音声データが混在している混在音声デ ータを、発言者毎の音声データに分離するステップと、 (2)発言者毎に該発言者を特定するための特定パラメ ータを準備するステップと、(3)分離された前記発言 者毎の音声データにつき、前記特定パラメータを参照し て、発言者を特定するステップとを有することを特徴と する発賞者特定方法である。このような第6の発明によ れば、例えば、会議の録音データなどに記録された、複 数発言者の音声や雑音などが含まれたの混在音声データ を音源ごとに分離し、各分離されたの音声データの発言 20 者を特定することによって、例えば、自動的に会議記録 データの作成などを行うことができる。

【0019】また、本出願に係る第7の発明は、第6の 発明である発言者特定方法において、前記特定パラメー タは、発営者が母音を発音した際のホルマント層波数で あり、分離された前記発言者毎の音声データにつき、ホ ルマント周波数を求め、求められたホルマント周波数に 関して、前記特定パラメータとしてのホルマント開波数 を参照して、発言者を特定することを特徴とする発言者 特定方法である。このような第7の発明によれば、フー リエ変換などの容易な処理で抽出できる特徴器であるホ ルマント周波数を用いて、各分離された音声データの発 言者特定を容易に行うことができる。

【0020】また、本出願に係る第8の発明は、第7の 発明である発言者特定方法において、前記特定パラメー タは、発言者が母音を発音した際の第1ホルマント周波 数及び第2ホルマント周波数であり、分離された前記発 営者毎の音声データにつき、第1ホルマント周波数及び 第2ホルマント周波数を求め、求められた第1ホルマン ト周波数及び第2ホルマント周波数に関して、前記特定 40 パラメータとしての第1ホルマント周波数及び第2ホル マント周波数を参照して、発言者を特定することを特徴 とする発言者特定方法である。このような第8の発明に よれば、第1と第2のスペクトルピークである2つのホ ルマント周波数を用いて発言者の特定を行うことによっ て 容易に、かつより正確に特定を行うことができる。 【0021】また、本出願に係る第9の発明は、第6の 発明乃至第8の発明のいずれかに記載の発言者特定方法 において、分離された前記発言者毎の音声データにつ き、前計特定パラメータを参照して発言者を特定するス 50 につき、複数の時点のホルマント周波数を求め、求めら

テップにて発言者を特定できなかった場合には、該音声 データから複数の時点のホルマント周波数を求め、求め られた複数時点のホルマント温波数に関して、前記特定 パラメータとしての複数時点のホルマント周波数を参照 して、発言者を特定することを特徴とする発言者特定方 法である。このような第9の発明によれば、ある音声の 発声者を特定する上での特徴量であるホルマント周波数 の、時間的変動をも考慮することにより、より正確に発 言者の特定を行うことができる。

【0022】また、本出願に係る第10の発明は、第6

10

の発明乃至第9の発明のいずれかに記載の発言者特定方 法において、分離された前記発言者毎の音声データにつ き、前記特定パラメータを参照して発言者を特定するス テップにて発言者を特定できなかった場合には、該音声 データから有声音データを分離し、該有声音データにつ き、ホルマント周波数を求め、求められたホルマント周 波数に関して、前記特定パラメータとしてのホルマント 間波数を参照して、発営者を特定することを特徴とする 発言者特定方法である。ホルマント周波数による発言者 の特定は、有声音、特に母音の識別に有効であるので、 このような第10の発明によれば、無声音を含む様々な 音声をもより正確に識別することができる。ここで、無 相関化処理及び独立成分分離処理がなされる前の音声デ ータが複数人の音声が混在しているデータであるのに対 して、無相関化処理及び独立成分分離処理という二つの 処理によって分離された分離音声データは、ある一人の 音声が抽出されたデータとなっている。よって、このよ

うな二つの処理によって分離された分離音声データから

は有声音を高い精度で抽出することができる。 【0023】また、本出願に係る第11の発明は、第1 30 ①の発明の発言者特定方法において、分離された前記発 言者毎の音声データにつき、前記特定パラメータを参照 して発言者を特定するステップにて発言者を特定できな かった場合には、該音声データから有声音データを分離 し、該有声音データにつき、第1ホルマント周波数及び 第2ホルマント郷波数を求め、求められた第1ホルマン ト周波数及び第2ホルマント周波数に関して、前記特定 パラメータとしての第1ホルマント周波数及び第2ホル マント周波数を参照して、発言者を特定することを特徴 とする発言者特定方法である。このような第11の発明 によれば、分離された有声音データに対して、第1と第 2のスペクトルピークである2つのホルマント層波数を 用いて発言者の特定を行うことによって、より正確に特 定を行うことができる。

【0024】また、本出願に係る第12の発明は、第1 ①の発明または第11の発明の発言者特定方法におい て、分離された前記発言者毎の音声データにつき、前記 特定パラメータを参照して発言者を特定するステップに て発言者を特定できなかった場合には、該有声音データ

れた複数語点のホルマント最適数に関して、前記縁定パ ラメータとしての複数語点のホルマント原接数を参照し て、発言者を特定することを特徴とする発言者特定方法 である。このような第12の発明によれば、分離された 有声音データに対して、発音者特定上の特徴量であるホ ルマント語波数の時期的変動をも考慮することにより、 より正確に発言者の特定を行うことができる。

[0025]また、本出額に係る第13の発明は、第1 ()の発明又は第11の発酵の発音者特定方法において、 前記音声データから前記有声音データを分離する際に、 該音声データに対して独立成分に分離するための独立成 分分離処理が行われることを特徴とする発言者特定方法 である。有声音は声帯の振動を伴うものなので、このよ うな第13の発明によれば、音声データに独立成分分離 処理をかけることによって、声帯の振動を伴わない無声 音と声帯の振動を伴う有声音とを容易に分離することが 可能となる。

【0026】また、本出類に係る第14の発明は、複数 発言者の音声データが混在している混在音声データか 発明乃至第13のいずれかの発明の発言者特定方法によ り、分離された前記発営着毎の音声データにつき、発管 者を経定するステップと、特定された発言者と 移発音 者の発言とを対応付けて紀線媒体に出力することによ り、議事録を作成するステップとを有することを特徴と する議事録作成方法である。このような第14の発明に よれば、発言者の特定が自動的に正確に行われるため、 長時間にわたる会議の議事録作成を自動的に行うことが でき便利である。

発言者の音音データが混在している混在音声データを、 発言者毎の音声データに分離する音声データ分離装置に おいて、前記混在音声データを互いに無相関化するため に無相關化処理を行い、前記無相關化処理の行われたデ 一タを独立成分に分離するために独立成分分離処理を行 うことを特徴とする音声分離装置である。このような第 15の発明によれば、入力される温存音声データ(生デ ータ) に含まれる各音声テータの相関性、及び独立性の 両性質をともに考慮し、複数の音声データや混入する雑 音などの有する相關性や独立性が 時間的・空間的に変 40 動する場合でも、発覚者能の音声に正確に分離すること が可能な音声分離装置を実現できるまた、本出願に係る 第16の発明は、第15の発明である音声分離装置にお いて、前配独立成分分離の行われたデータの分離性が不 十分な場合には、分離性が十分になるまで、前記独立成 分分離処理の行われたデータについて、前記無料関化処 理及び前記独立成分分離処理を繰り返し行うことを特徴 とする音声分離装置である。このような第16の発明に よれば、混在音声データを音源毎の音声データに充分に 分離させることの可能な音声分離装置を実現できる。

[0028] 全た、本出願に係る第17の類似は、第1 5 又は第16の発明である音度分離装置において、デー タの件質により、前記独立成分分類処理として、非ガウ ス性のデータを独立成分に分離するための非ガウス性独 立成分分離処理、非定常性のデータを独立成分に分離す るための非定常性独立成分分離処理。有色性のデータを 独立成分に分離するための有色性独立成分分離処理、の うちのいずれかの処理を行うことを特徴とする音声分離 装置である。このような第17の発明によれば、無相関 10 化処理の行われたデータの性質に応じて最適な独立成分 分離処理を行うことができるから、混在音声データを音 酒毎の高密データにより効果的に分離させることの可能

【0029】また、本出線に係る第18の発明は、第1 7の発明である音声分離装置において、最初に行われる 独立成分分離処理は、非ガウス性のデータを独立成分に 分離するための非ガウス性独立成分分離処理であること を特徴とする音声分離装置である。非ガウス性独立成分 分離処理は他の独立成分分観理方法に比べてその前処理 ら、緑事録を作成する繊華録作成方法において、第6の 20 としての無相關化処理の影響を受けやすいから、このよ うな第18の発明によれば、最初に非ガウス性独立成分 分離処理を行うことにより、無相関化処理がうまく実行 されたかどうかを、該舞網際化処理に引き続く非ガウス 性独立成分分離処理によって効果的に評価することが可 能な音声分離装置を実現できる。

な音商分離装置を実現できる。

【0030】また、本出額に係る第19の発明は、第1 5 乃至第18の発眼である音声分離装置において、前記 無相関化処理は、少なくとも主成分分析及び因子分析を 行うことを特徴とする音声分離装置である。このような 【0027】また、本出郷に係る第15の発明は、複数 30 第19の発明によれば、各主成分の寄与率を求めて累積 容与率が所定のしきい値を越えるところの成分数を次数 とすることなどにより、採用する主成分データの数(次 数) を決定した上で、効果的に無相関化処理を行うこと が可能な音声分離装置を容取できる。

> 【0031】また、本出額に係る第20の発明は、複数 発言者の音声データが顕在している誕在音声データを、 発営者毎の音両データに分離し、波発言者毎の音声デー タにつき発言者を特定する発言者特定装置において、第 15万至第19のいずれかの発明の音声分離装置によ り、複数発言者の音声データが派だしている源在音像デ ータを、発言者毎の音声データに分離し、分離された前 記録討者毎の音声データにつき、発言者毎に終発言者を 特定するための特定パラメータを参照して発言者を特定 することを特徴とする発音者特定装置である。このよう な第20の発明によれば、例えば、会議の録音データな どに記録された、複数発言者の音声や雑音などが含まれ たの混在音声データを音額ごとに分離し、各分離された の音声データの発言者を特定することによって、例え ば 自動的に会議記録データの作成などを行うことの可 50 能な発言者特定装置が実現できる。

【9032】また、本出版に係る第21の発明は、第2 のの預明である完全者特定接額において、前記特定パラ メータは、完全者が持ちを発した使のホルマント間被 数であり、分離された前記性含著毎の資産データにつ き、ホルマント買該数を求め、求められたホルマント間 減数に職して、施制を記して、少らとしてのホルマント 間波数を参照して、発言者を特定することを特徴とする 発言者特定装置である。このような第21の発明によれ は、アーリエ変をとどの容易と処理で始化できる特徴 であるホルマント同波数を用いて、各分離された音声デ 10 ータの復言者特定を容易に行うことの可能な発音者特定 数数が実現である。

[00年3] また、本出版に係る第22の契則は、第2 1の発明である発言者特定装置において、前配特にバラ メータは、発言者が母音を発症した際の第1ネルマント 周波数及び第2ネルマント周波数であり、分離された前 社発言者等の音声データにつき、第1ネルマント周波数 及び第2ネルマント周波数を示め、求められた第1ネル マント周波数なが第2ネルマント周波数数の第2 ネルマント周波数を参照して、第言者を特定することを 特徴にガラメータとしての第1ネルマント周波数数をが第2 ネルマント周波数を参照して、発言者を特定することを 特徴とする発言者特定装置である。このような第22の 部形におはば、第1と第2のスペクトルピータである2 つのホルマント周波数を用いて発言者の特定を行うこと によって、各場に、かつより正確に特定を行うことの可 能な好き事態が実際が実際である。

【0034】また、本出版に係る第23の発明は、第2 のの発明乃至第22の発明のいずれかに記載の発言者特 定要域において、分離された応記澄言者様の音声データ につき、前記特定パラメータを参照して発言者を特定で なかった場合には、該音曲アータから複数の場点のホ ルマント周波数を求め、求められた複数時点のホ ルマント周波数を求め、求められた複数時点の赤 ルマント周波数を求め、求められた複数時点の赤 ルマント周波数を剥め、ボウムーをとしての複数時 点のホルマント履数数を参照して、発音者を特定すること とを特数する存置者特定数据である。このよう左第2 3の発明によれば、ある音声の発声者を特定する上での 特徴官であるネルマント魔後数の。期間的変動をも考慮 することにより、より正確に発言者の特定を行うことの 可能な写信者特定認か実現できる。

よれば、無声高を含む様々な音声をもより正確に鑑例することができる。こでで、無相關化処理及び独立成分分 機処理がなよれる前の音声子・少が複数人の音が現在しているデータであるのに対して、無相関化処理及び独立成分分離処理という一つの処理によって分離されたデータとなっている。よって、このような二つの処理によって分離された分離後度データからは有声音を高い精度で抽出することができる。

14

- 【0036】また、本比額に保る第25の2等別は、第2 4の発明の発音者特定設價において、分隔された前記発 言者4の脅声データにつき、前記特定パラメータを参照 して発き者を特定できなかった場合には、感疹相データ から有声音データを分離し、液育声音データとつき、第 1ホルマント高波数及び第2ホルマント周波数を求め、 求められた第1ホルマント周波数及び第2ホルマント周 被数に関して、前記特定パラムータとしての面1ホルマ ント周波数及び第2ホルマント周波数を参照して、発音 番を特定することを特徴とする発音が完定展置である。 でのような第25の発明によれば、分離されて寿命音デ
- このような第25の発明によれば、分離された有声音データに対して、第1と第2のスペウトルピークである2つのホルマント園波数を用いて発言者の特定を行うことによって、より正確に特定を行うことの可能な発言者特定変置が実現できる。

【0037】また、本出額に係る第26の発明は、第24の預明または第25の発明の発音者特定接臘よおい、分離された前記発音番毎の資声データにつき、前記 特定パラメータを参照して発音帯を特定できなかった場合には、該有声音データにつき、複数の時点のホルマン

ト脳級数を求め、求められた機数項点のホルマント周波 数に関して、前記特定パラメークとしての複数時点のホ ルマント開放数を参照して、発音者を特定することを特 後とせる発音者検定級限である。このような第26の発 明によれば、今機とれた青田子データに対して、発言者 特定上の特徴量であるホルマント周波数の時間的変動を も着渡することにより、よりに確じ得着の対定を行う ことの可能な作者を対象と数置が実現できる。

【0038】また、本出類に係る第27の発明は、第2 4の発明又は第25の発明の発置者特定装置において、

約 前記音用データから前記有声音データを分離する際に、 前記音用データに対して独立成分に分離するため、独立成分分離処理が行われることを特徴とする発言者や定襲である。有声音は声部と無動を作うこのなので、このような第2つが発揮によれば、音声データは地を分分離処理をかけることによって、声帯の推動を伴わない無声音と声帯の振動を伴う有声音とを容易に分離することの可能な代書者特定変勝が実施となる。

【0039】また、本出額に係る第28の発明は、複数 発言者の音声データが配在している混在音声データか を、要素品を作成する発素等作成装置において、第20

音の離別に有効であるので、このような第24の発明に 50 ら、議事録を作成する議事録作成装置において、第20

(9)

乃至第27のいずれかの発明の発言者特定装置により。 分離された前記発言者毎の音声データにつき、発言者を 特定し、特定された発言者と、該発言者の発言とを対応 付けて記録媒体に出力することにより、議事録を作成す ることを特徴とする議事録作成装置である。このような 第28の発明によれば、発言者の特定が自動的に正確に 行われるため、長時間にわたる会議の議事録作成を自動 的に行うことの可能な議事録作成装置が実現できる。

【0040】また、第1乃至第5のいずれかの発明の音 声分離方法を音声分離装置に実行させるためのコンピュ 10 が、まず無相関化処理W 1 に入力される。 ータプログラムも実現可能である。

【0041】また、第6万至第13のいずれかの発明の 発言者特定方法を発言者特定装置に実行させるためのコ ンピュータプログラムも実現可能である。

【0042】また、そのようなコンピュータプログラム を記録したコンピュータ読み取り可能な記録媒体も実現 可能である。

[0043]

【発明の実施の形態】 == 混在音声データの音声分離=

以下、図面を参照しつつ、本発明のより具体的な実施形 態につき、詳細に説明する。まず、本発明の方法の前半 部分である、混在音声データの音声分離ステップについ て説明する。

【0044】本実施形態では、2人で行われたある会議 の発言内容の音声データを2本のマイク(マイク1、マ イク2) で拾う。図1は、そのうちマイク1から入力さ れた音声データ(生データ) Xの波形である。この源在 音声データには、複数の発言者の音声データが混在して いてもよい。2人の発声をそれぞれ音順5.1. 5.2と呼 ぶことにする。

【0045】図2は、音声分離処理のサイクルを示す図 である。マイク1及びマイク2から入力された混在音声 データは、まず無相関化処理W1にかけられる。無相関 化処理W1に渡される音声データは、図1の[1] [2] のようにセグメント化されて1つずつ渡される。最も効 率がよいように、各セグメントは互いに1/2周期ずつ オーバーラップしている。

【0046】図2において、無相関化処理W1の次のス 40 テップである I C チューナーは、独立成分解析 (I C A) の手法を3種類のうちから選択するためのチューナ ーである。その次のステップである独立成分分離処理W 2は、非ガウス性に基づく分離処理W2(a)、非定常 性に基づく分離処理W2(8)、有色性に基づく分離処 理W(v)の3種類のうちいずれかの方式の処理を行 う、W2の後のステップの評価器Eでは、W2にて分離 されたデータの分離性の評価を行う。マイクから入力さ れた混在音声データの音声分離性能が充分になるまで、 以上のW 1 → 1 C チューナー→W 2 → E というサイクル 50 重要た点である。m番目の図子までを採用する場合を、

を繰り返し回す。ただし、1回目のサイクルでは、独立 成分分離処理W2として、非ガウス性に基づく独立成分 分離処理W2(α)を行い、2回目以降のサイクルで は、ICチューナの選択に従って、W2(a)、W2

(β)、W2 (γ) の3種類のうちから適切な方式の独 立成分分離処理を行う。

【0047】図3は、1回目の音声分離サイクルを示し ている。図1における前記[1]の時間セグメントの、マ イク1及びマイク2からの混在音音データx1、x2

【0048】図7及び図8は、それぞれx1及びx2の デジタル化波形図データ(縦軸は音の強さで、単位はミ リポルト) を示す。各時点のx1、x2データを、横軸 をx1の強さ、縦軸をx2の強さとして散布図を描くと 図9のようになる。散布図は、第1象限から第3象限に かけて若干直線的な分布を呈し、x1とx2のデータは 互いに相関性を有することを示している。これら生デー タであるx1、x2が無相関化処理W1にかけられる と、互いに相関性を有しないデータ f 1、 f 2 に変換さ 20 れる。

【0049】f1及びf2の散布図を図10に示す。図 10の横軸は因子得点Fの第1因子f1、縦軸は因子得 点Fの第2因子f2を示している。図9が軸に対してい びつな平行四辺形状に分布していたのに対し、軸に対し でまっすぐで形の整ったひし形状に分布しており、f-1 とf2はもはや互いに相関性を有していないことがわか 3.

【0050】ここで、無相関化処理の内容について説明 する。図6は、無相関化処理W1の…例のフローチャー いるのみならず、音楽や、さらには雑音などが混ざって 30 トを示したものである。まず、図7及び図8に示した音 声生データx1、x2を(1)式により標準化する。標 準化の結果、平均が0、標準偏差1のデータとなる。 【数1】

$$\chi = \frac{(\tilde{\chi} - \mu)}{\sigma}$$
(1)

【0051】生データx1、x2の相関行列(ベクトル C)を(2)式より求める。(2)式において(x1、 x 2) はベクトルの内積を表す。 【数2】

$$\mathbb{C} = \frac{1}{1!} \left(X_1, X_2 \right) \tag{2}$$

【0052】上記相關行列に対する間有値 11と固有べ クトルAを(3)より求める。 [0053]

【数3】

(4)

7

m次元と呼ぶ。先に求めた顕有ベクトルAにより、 14) 式によって主成分2が栄まる。

[数4] ※=A×

【0055】次にm側の超子に対して、(5)式の形の 定義式にて因子分析を実行する。(5)式における e は、特殊因子と呼ばれるものである。

[数5] X=A聚+e (5)

【005 6】この房子モデルが(6)式の表現をとる。 (6)式における房子負荷量む1 J、因子利点Fは、 (7)式及び 18)式によって求める。そして、図6の フローチャートの最終ステップで、結局音声生データ は、互いに無相関な房子特点(ペクトルド)に変換される。

[数6] X=bF+e (数7]

 $\mathbf{b}\mathbf{i} = \sqrt{\lambda} \mathbf{j} \mathbf{A} \mathbf{i}$ (7)

[% 8] (8)

【0057】以上熱明したWIの主次持機は、主成分分析と関子分析とを組み合わせている点である。その効果は、主成分分析を実行すると各主成分の寄与事を同時に求めることができるので、例えば、第1次主成分から形成大主成分での服積等与480%を担けることにある。分離すべき音声生データは、時間的姿勢が大きく変行はよる相関の度化・が大きく変行するので、何個の因子を採用するかは無相関化処理において重要なしてある。

[0068] 発記者の人数があらかじめ判明している場合には、次数mを発話者の人数に固定してしまえばよいが、人数が不明なときは、例えば、異情密与率が所定のしない値を超またときの主張分数を実施します。次数mの決定方法は、システムに応じて様々な方法を準備しておき、防機疾薬に変化させる・チューニングする)ことが背よしい、次にこのチューニングの一実施例について終しく説明されている。

[0059] 図23は、システムに成じた方法で次数の を決定する手順を示すフローチャートである。図23 で、RKのは累積密与率の初期しさい値、Mは採用し得 る最大次数(次数の上側しさい値)、△RKは累積寄与 率の変化量である。主意分析を実行すると、図21の ような、表質の「深血主動なで採用したということを 示す)とその累積寄与率との関係を示すグラフが得られ る。図21にはA、B、C3種類のブラフの例を描いて いる。

【0060】まず、第1の処理ステップとして、累積寄 50 一における処理が加わっている。独立成分分離処理W2

与率R K にしき、億R K O (この実施制では8 0 %) を 設定しておき、このしきい確R K のを超える次数の表で、 める。ところが、実数がまりに大きいとその後の処理 が関縁に過ぎるので、あらかじめ次数の上限値制を決め てをこ。図2 1 の例では、M ー 1 ピラモと、 A の場合は しきい確R K O を超える次数m ー 2 であるので、 m ー 2 く4 = Mとなって、実数mは2 に決定される。 B の例で は R K O を超える実数mは5 であるので、 m ー 5 > 4 ー M となってしまい、実数mはまだ決定されない。 C の例 でも同様に来動は状定されない。 C の例 10 でも同様に来動は状定されない。 C の例

18

【0061】そのような場合は図22に示す、第2のステップを実行する。すなわち、次数mの増加に対する。 尽い の表現である。すなわち、次数mの増加に対する。 K の差分変化電点 R K を調べる。これは要するに、累 稼事与率の変化が最大となる沙数mをもって採用すべ、き 数数とするという処理方法である。この実施物では、き め間ではm=2、この例ではm=4においてんR K か識 大値をとる。この場合は次数mが上膜値Mよりも下なら ば、その次数mを採用とするが、Mを上回る場合は、そ の処理が次のステップに送るれる。

【0063】突動が水ド K n = 50%以上で、かつM以 下の値で発見されない場合は、再び上記第20ステップ と同様の処理・すなわら入R K が販大になる次数を求め て、その値を次数加として採用してしまう。これは、異 報寄与事が大きく変化するということは、その次数の前 後で情報がより多く保存されるということを流敷やするの で、少なくともその次数までは採用したい、という考え に掻づくものである。

【9964】以上のようにして 図3において、無相関 佐されたデータ「1、「2は、ただちに独立双分分離処 地 選収2に送られる。」毎日の音声分離サイクルでは、こ れらの無相関化チータ「1、【2に対し、非ガウス性に

基づく独立成分分離処理W2(a)を実行する。

【0065】以上、図3におけるW1及びW2(a)の処理により、分離信号 aまよびbが得られ、これらの分離性、代力な分類ない。 分離が不十分なとる(図のま1)はこれらる、bのデータに対して、2回目のサイクルを実行さる。

【0066】2回目のサイクルの例を図4に示す。図3 に示した1回目のサイクルと似ているが、ICチューナ 19 を行う前に、1 (チューナーで 2 回目の無相関化処理さ れたデータ I 1 、 f 2 「の信号特性を解析し、非ガウ ス性に基づく処理W 2 (a)、非定常性に基づく処理W 2 (身)、有色性に基づく処理W 2 (y)のかずれをW 2 と して実行するかを選択する。この例ではW 2 (身)を実 20 された入力データのいずれもが有色性を有すると評価された基合、ICチューナーは、独立成分分離処W2として有色性に基づく処理W2(y)を遊訳する。

【0071】 例5は3回目のサイクルを示している。各 処理は2回目のサイクルと同様であるが、3回目の独立 ペンパンが加まり、マスポーラントをもみに繋っている。と